未折叠蛋白质的普遍初始热力学亚稳态

马晓亮[1] 侯成宇[2] 史丽萍[1] 李隆[3] 李佳成[1] 叶林[4] 杨霖[1,4*] 赫晓东[1]

[1]（哈尔滨工业大学特种环境复合材料国防重点实验室 哈尔滨 150001 中国）
[2]（哈尔滨工业大学电子与信息工程学院 哈尔滨 150001 中国）
[3]（哈尔滨理工大学电气与电子工程学院 哈尔滨 150080 中国）
[4]（悉尼大学机械与机电工程学院 悉尼 2006 澳大利亚）

摘要 探索和理解蛋白质折叠问题一直是分子生物学的挑战。未折叠的蛋白质应该存在一种普遍初始热力学亚稳态，否则无法解释蛋白质是如何在剧烈的热振动干扰下完成精确折叠的。文章通过分析水溶液环境和蛋白质折叠的相关性，揭示了一种由水分子屏蔽效应引起的未折叠蛋白质的普遍初始热力学亚稳态，该亚稳态的存在是环境中水分子的物理性质决定的并被认为是蛋白质正确折叠的必要条件。我们通过研究已发表的实验数据和建立分子模型，找到了该普遍初始热力学亚稳态存在的相关证据，通过理论分析初步探索了该热力学亚稳态导致蛋白质精确折叠的相关物理机制。

关键词 蛋白质折叠 热力学亚稳态 水分子屏蔽效应

学科分类号 Q7

1 引言

蛋白质及其产物是地球生命的基础，几乎所有已知的重要生物化学反应和生命现象都是由蛋白质参与完成的[1]。地球生物中可能存在数百万种不同的蛋白质，每种蛋白质的生物学功能和活性都是通过其三维形状表达的，而蛋白质能够形成三维结构的原因是蛋白质的折叠，因此蛋白质折叠可以被看作是分子生物学的基础[2,3]。未折叠的蛋白质也就是多肽链，其往往不具有生物学功能的原因是多肽链本质上只是类似树脂分子的长链聚合物分子。蛋白质折叠实现了多肽链的功能化，例如，蛋白质折叠形成了球状蛋白质复杂的表面活性。蛋白质折叠可以被看作是一切生物存在、进化、功能化和多样性的最重要的机制和原动力，并直接导致了不同种类蛋白质的产生[4]。自从Anfinsen 揭示了蛋白质的三维结构由蛋白质的一级结构决定这一重要关系，探索蛋白质氨基酸序列和蛋白质三维结构间物理关系的研究成了分子生物学的重大前沿领域[5]。蛋白质折叠问题在 60 年前被提出，作为最重要的分子生物学问题对蛋白质折叠机理和折叠规律的探索被广泛开展，相关研究多次获得诺贝尔化学奖[5-8]。但是该问题在生物物理层面的研究进展依然有限，至今仍无法解释自然是如何完成和确保蛋白质精确折叠的，这限制了蛋白质研究的发展和应用。蛋白质折叠问题被认为是基础科学的遗产，蛋白质折叠问题的解决将会提高分子生物学和药物研发的水平和进程，其影响将会广泛而深入[1]。Ken 将蛋白质折叠问题细化为不同层面的三个经典问题[1]：(i) 存在于蛋白质氨基酸序列中的决定蛋白质三维结构的物理折叠密码是什么？(ii) 导致蛋白质快速折叠的物理驱动力和机制是什么？(iii) 是否能通过计算机算法准确预测蛋白质的三维结构？蛋白质的折叠过程主要由如下的多种物理力所引导：(i) 形成氢键、(ii) 范德华力、 (iii) 静电力、(iv) 疏水作用和(v) 熵[1]。这些物理作用力可以近似地由“力场函数”描述，这使得分子动力学模拟方法成为了解决蛋白质折叠问题的重要手

* 通讯作者. E-mail: linyang@hit.edu.cn

段之一。然而即便是使用现今最高性能的计算机进行模拟，蛋白质折叠问题依然没有被解答。

蛋白质分子的折叠主要导致其分子内部形成大量的氢键和疏水键，这些氢键和疏水键的形成过程释放了大量的自由能并降低了分子结构的势能，所以或许可以将蛋白质的折叠过程看作是分子结构的自由能释放过程或分子结构松弛过程[9]。氢键是蛋白质结构稳定性的最主要贡献者，值得指出的是温度的增加会破坏蛋白质三维结构的稳定，并使蛋白质发生不可逆转的变性，变性后的蛋白质结构可能也是一个稳定的折叠构象。绝大多数的蛋白质只能在特定的温度范围内才能发生正确折叠的原因是未知的。

一个蛋白质拥有巨量的潜在可能的折叠构象， 由于这些潜在可能的折叠构象往往都形成了大量的氢键和范德华力，所以每一种折叠构象的自由能都远小于未折叠的蛋白质的自由能。如果把蛋白质折叠过程看成是简单的自由能释放过程，那么蛋白质如何能快速地折叠成其唯一的天然结构而不发生错误折叠[10]？考虑到一些蛋白质的折叠过程只需要几微秒，蛋白质是如何能不通过搜索折叠路径就避免了错误折叠的？对任意氨基酸序列的折叠途径还没有定量的微观理解。

激光温度跳跃技术和单分子实验技术等实验手段已经揭示了蛋白质折叠动力学的一些重要规律。实验结果已经证实了蛋白质折叠首先导致二级结构的形成，蛋白质的三级结构的形成晚于蛋白质的二级结构[2]。此外实验证实了蛋白质二级结构形成的速度非常快，可以在微秒尺度内完成二级结构的折叠[11,12]。蛋白质的天然结构往往被认为是其自由能最小的最稳定的结构[13]，因此蛋白质的二级结构可以被推测为势能最小的局部最稳定结构。经典的二级结构α螺旋和β折叠确实极大化了主链间生成氢键的数量，因为主链上的羰基氧原子（C-O）和酰胺氢原子（N-H）都彼此生成了氢键[14,15]，这些大量氢键的生成确保了α螺旋和β折叠的形成过程可以有效的释放自由能和降低分子结构的势能。

但是大量蛋白质的二级结构并不是α螺旋和β折叠结构[14,15]，也就是说许多蛋白质的二级结构并没有有效地形成更多数量的氢键来大幅减小其自由能，所以说这些二级结构并不是局部分子结构势能最小的稳定结构。也就意味着这些非典型的二级结构的稳定性并不高，因为维持这些二级结构稳定的氢键的数量并不大，其结构稳定性明显无法和α螺旋和β折叠结构比。既然蛋白质折叠的自然过程无法保证蛋白质氨基酸序列中的每个片段都形成势能最小化的稳定结构，那么蛋白质折叠过程就不能被简单的理解为松弛过程或自由能最小化过程。因为很难解释蛋白质折叠是如何通过放弃了大量的局部更稳定的结构来追求三级结构的稳定性和三级结构分子势能的最小化。蛋白质的疏水内核的形成是发生在蛋白质折叠的后期，疏水键的形成是需要疏水氨基酸相互靠近到疏水区域的表面张力相互作用的距离，因此不能确定是疏水作用主导了蛋白质的折叠过程。综上，蛋白质折叠导致不同氨基酸序列片段形成不同的二级结构是不能简单地解释为普通的自由能释放现象。是什么机制导致蛋白质在特定的氨基酸序列片段形成特定的二级结构是未知的，也就是蛋白质的物理折叠密码是未知的。

此外蛋白质折叠往往需要水溶液环境和适宜的温度，蛋白质折叠所需的环境温度接近室温，该温度带来的热运动对于纳米尺度的未折叠蛋白质分子来说是非常剧烈的，甚至可以认为未折叠的蛋白质分子是在剧烈热振动中完成的精确折叠。在热运动如此剧烈的环境中，蛋白质精准的折叠几乎不可能是完全由结构松弛过程或疏水作用所决定的。因为未折叠的蛋白质分子表面存在大量的亲水官能团，剧烈的随机热振动会导致这些亲水官能团随机相遇形成氢键并由此导致错误的折叠。由此可以推测未折叠的蛋白质分子可能存在某种热力学亚稳态，这种亚稳态可以避免蛋白质在振动中发生随机的折叠，蛋白质的折叠可能是利用热振动来精确破坏某些氨基酸序列片段的这种热

力学亚稳态来形成精确的折叠，这也解释了恰当的温度对蛋白质折叠的必要性。因此，寻找未折叠的蛋白质的热力学亚稳态是解答蛋白质折叠问题的关键。

如果所有未折叠的蛋白质都普遍存在同一种热力学亚稳态来保证蛋白质能按清晰的物理折叠密码折叠，那么该亚稳态很可能是在蛋白质折叠发生前就存在了，也就是说多肽链分子本身就具有某种热力学稳定状态。因为一旦蛋白质开始折叠，不同种类的蛋白质在折叠过程中的很难形成类型相同的某种亚稳态。

值得指出的是，蛋白质总是在水溶液环境中才能折叠，蛋白质几乎不能在任何非水溶液中正确折叠[16, 17]，全面的理解未折叠的蛋白质与水的相互作用可能会导致发现蛋白质折叠问题的答案，或至少是发现蛋白质折叠路径上重要的热力学亚稳态。相比于其他的溶剂分子来说水分子有其独特性，水分子的尺寸是非常小的，且每个水分子的氢原子和氧原子分别带有带正电和负电，这就使得水分子可以插入到未折叠蛋白质的侧链之间的空隙中，并和未折叠蛋白质的每个亲水侧链形成氢键，也就是说这些亲水侧链生成氢键的能力可以被大量的水分子饱和，这就避免了亲水侧链之间生成氢键的可能，亲水侧链和主链之间也不会形成氢键。水溶液中的水分子可以屏蔽每个亲水侧链的亲水性，在水溶液环境导致亲水侧链失去活性的情况下，未折叠蛋白质的状态很可能是由蛋白质主链上其他的相邻带电原子决定的。

2 结果

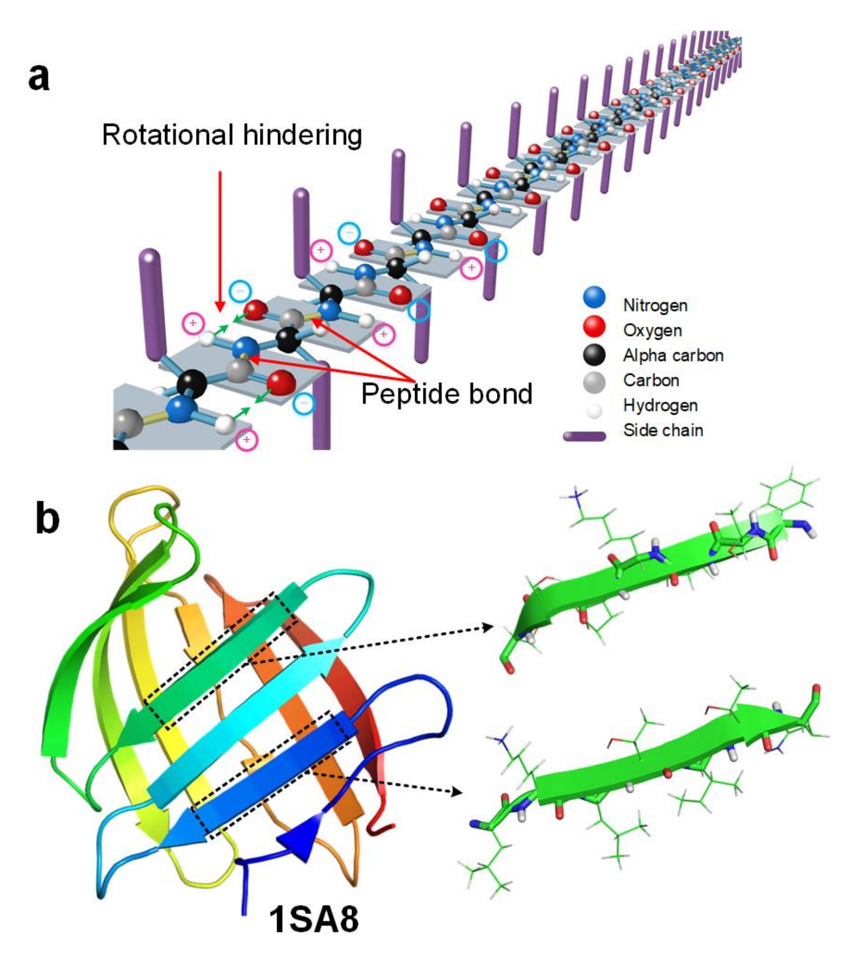


图 1 未折叠蛋白质的热力学亚稳态

a. 水环境中多肽链相邻肽平面保持平行的稳定状态　b. 相邻肽平面保持平行的状态广泛存在于β折叠中

在水环境中，一个未折叠蛋白质的构象可能首先进入一个热力学亚稳态来避免蛋白质发生错误折叠，这个亚稳态可能也是蛋白质物理折叠密码生效的必要条件，如图 1a 所示。蛋白质主链上的肽键由于有双键属性所以无法自由旋转。每个肽平面的羰基氧原子和相邻肽平面内的酰胺氢原子分别带有正电和负电，水分子无法屏蔽相邻如此近的两个原子间的静电吸引。这种羰基氧原子和酰胺氢原子间的正负电吸引会导致相应的 C-O 键和 N-H 键趋于平行（见图 1a），这同时会导致 C-O 键和 N-H 键所在的两个相邻的肽平面趋于平行，考虑到肽平面是刚性平面结构，我们会发现这种状态下的未折叠的蛋白质主链是无法自由折叠的，因为其所有的单键的旋转都受到了限制。如果没有热运动破坏这种羰基氧和酰胺氢原子间的静电吸引，那么不可能发生蛋白质的折叠。只有相邻的羰基氧和酰胺氢原子远离彼此并摆脱了静电吸引，主链上的化学键才能自由扭转。由此可见相邻的肽平面保持平行的状态就是未折叠的蛋白质的一种普遍的热力学亚稳态。存在这种热力学亚稳态的最直接的证据是β折叠的分子结构状态就是相邻的肽平面保持平行，这意味着该亚稳态大量的存在于蛋白质天然结构中，如图 1b 所示。此外即使在非β折叠的蛋白质天然结构中，这种相邻肽平面保持平行的状态也是普遍存在（如图 2 所示）。这证明了该相邻肽平面保持平行的状态对于多肽链来说是一种重要的热力学稳定状态。

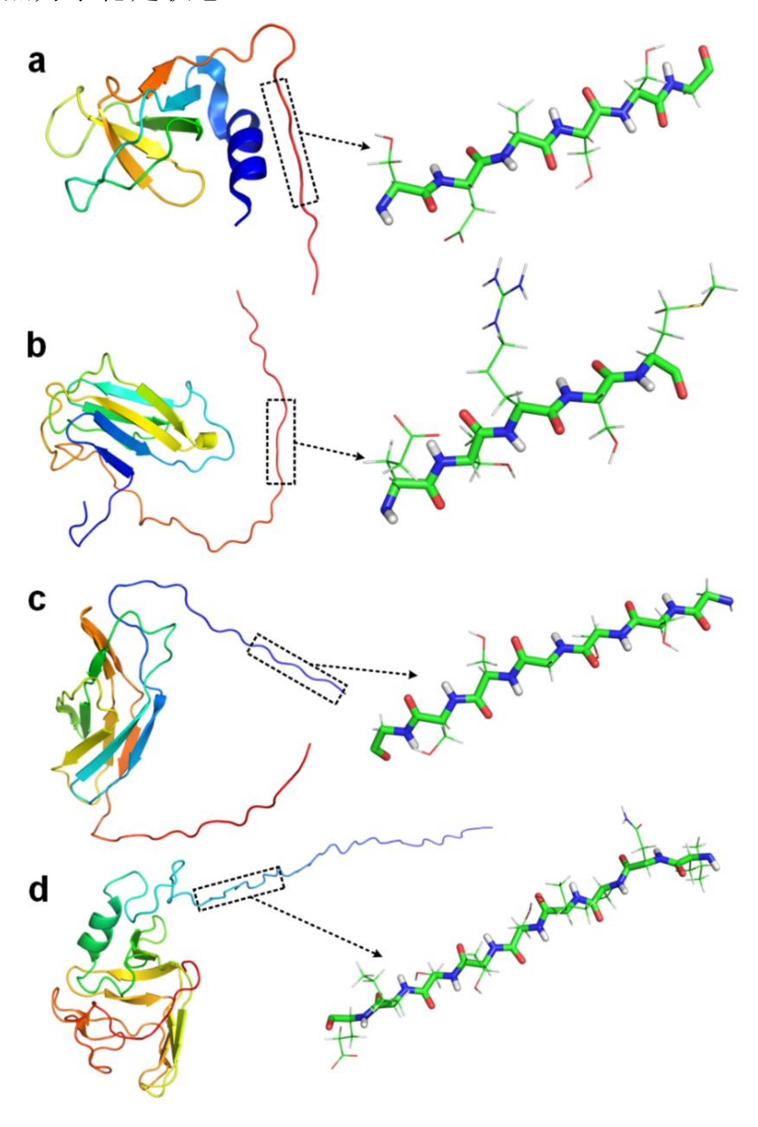


图 2 相邻肽平面保持平行的状态广泛存在于蛋白质的其他二级结构中

a. 1UG1　b. 1WFI　c. 1X5H　d. 1YDU.

有证据可以证明水分子确实屏蔽了亲水侧链与其他亲水侧链或主链之间形成氢键。首先在经典 CHARMM[18]势函数描述的蛋白质分子模型中，水分子中的氢原子带电量为（0.417e），其电量普遍大于蛋白质亲水侧链带正电原子的带电量（如图 3 所示）。所以在水环境中，蛋白质亲水侧链中的电负性原子只会和水分子中的氢原子先形成氢键。基于同样的原因，蛋白质亲水侧链中的带正电的氢原子只会和水分子中的带负电的氧原子先形成氢键。另一个证据是很多的蛋白质天然结构并不是紧实的[14,15]，很多局部结构之间还是可以靠近生成大量的氢键形成更为紧密的更为稳定的蛋白质结构，这些不紧实的蛋白质结构的存在也证实了水分子屏蔽亲水侧链的效果（如图 2 所示）。此外，蛋白质二级结构是通过主链间形成氢键来稳定的，亲水侧链没有参与蛋白质二级结构的形成，这也证明蛋白质的亲水侧链被水分子屏蔽了。

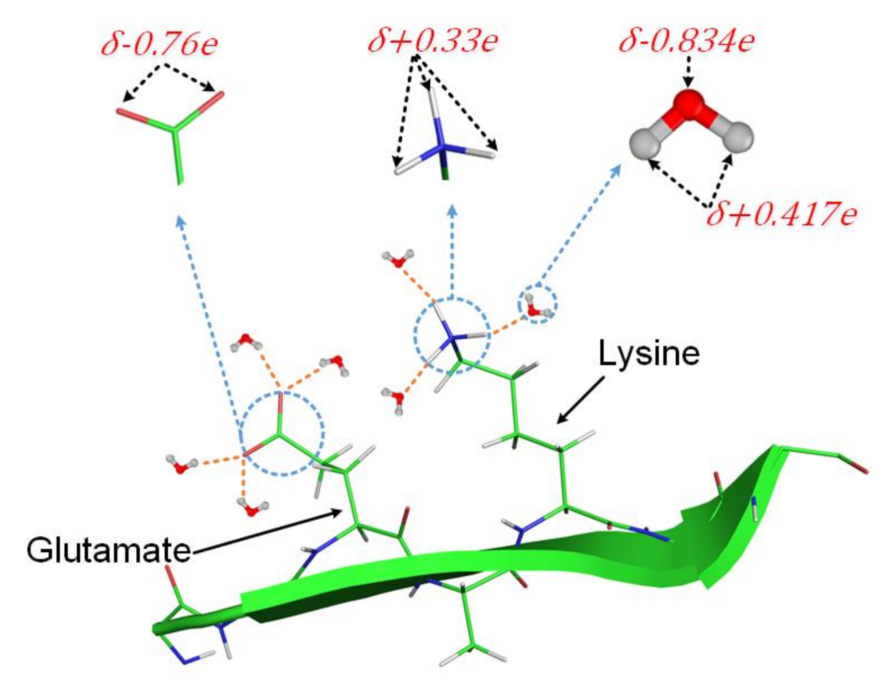


图 3 水分子的带电性与亲水侧链带电性的比较

考虑到在核糖体生成蛋白质氨基酸序列的过程中，游离的氨基酸分子中的羰基氧原子和酰胺氢原子已经发生了静电吸音，所以该亚稳态是伴随蛋白质氨基酸序列生成而立刻存在的。从 PDB 分子模型中可以发现 β 折叠中的相邻的羰基氧原子和酰胺氢原子的距离约为 2.2Å，羰基氧原子和酰胺氢原子在这个距离上的相互作用已属于弱氢键连接，这说明该热力学亚稳态是很稳定的。该亚稳态中相邻的肽平面间的这种弱氢键连接可以让未折叠的蛋白质主链可以承载一定的扭矩和弯矩，一个肽平面的旋转会带动相邻的肽平面也发生旋转，也就是说扭转波会沿着未折叠蛋白质的主链传播。

由于蛋白质折叠是需要一定的温度范围的，可以推测是温度引起的热运动破坏了序列中特定氨基酸上的亚稳态，也就是导致了该氨基酸的羰基氧原子和酰胺氢原子间的距离拉开到了相对应的 C-O 键和 N-H 键的平行状态无法恢复的转台，这就导致了该氨基酸所在的位置发生了折叠。那么这种亚稳态下的未折叠的蛋白质是如何准确折叠的？这要分析是什么原因会导致羰基氧原子和酰胺氢原子间的静电吸引被破坏。考虑到热环境必然会引起蛋白质的侧链振动，最有可能的原因就是相邻肽平面间与 α 碳原子连接的侧链发生了较大的振动或转动，破坏了相应的羰基氧原子和酰胺氢原子间的静电吸引状态。

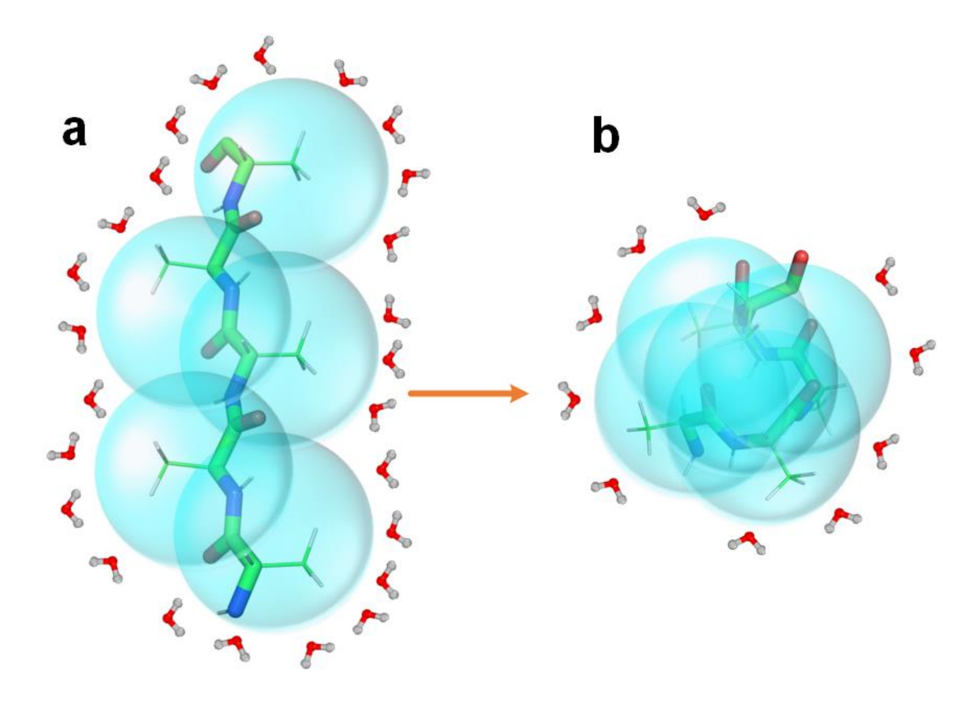


图 4 疏水塌缩导致失稳和α 螺旋生成

温度是通过什么具体的物理机制破坏了未折叠的蛋白质上某个氨基酸上的羰基氧原子和酰胺氢原子间的静电吸引状态从而启动折叠的？相邻的肽平面保持平行的亚稳态会导致每个肽平面都不能单独自由旋转，所以由侧链热振动产生的扭转波会沿着未折叠蛋白质的主链传播。当蛋白质氨基酸序列中相邻侧链的扭转阻力差异阻碍了扭转波沿蛋白质主链的传递时，就很可能导致折叠的发生[19]。温度引起的扭转波沿主链传播过程中可能还会启动相邻侧链间的静电吸引或静电排斥，这也很可能导致蛋白质折叠的发生。此外当多个疏水氨基酸相连的片段存在于未折叠蛋白质的氨基酸序列中，连续的疏水氨基酸很可能形成了贯通的疏水区域。在熵的作用下，这个不规则的疏水区域存在使其塌缩的疏水作用，疏水作用和侧链热振动配合也很可能破坏这几个氨基酸的热力学亚稳态。当这个片段的氨基酸失去了亚稳态后其疏水区域的塌缩很可能将这几个氨基酸形成了螺旋结构，如图 4 所示。如果可以发现连续 5 个疏水氨基酸所组成的折叠码广泛存在于蛋白质结构的 α 螺旋中，我们就可以确定其为蛋白质的物理折叠密码。通过对比 PDB 文件[20]，我们共查找了 30 个这样的码，发现 90%的码都存在于 α 螺旋中，证明了该折叠码的存在。

3 结论

未折叠蛋白质主链上可旋转的单键使得氨基酸序列中每个氨基酸都至少有两个折叠的自由度，这导致了蛋白质有无数种可能的折叠构象。考虑到水溶液中蛋白质折叠所需的温度环境会使未折叠蛋白质产生剧烈的热振动和错误的折叠，可以推测存在一种普遍初始热力学亚稳态来避免蛋白质在折叠过程中受到热振动的干扰。通过发现水分子具有屏蔽亲水侧链亲水性的能力，我们揭示了以相邻肽平面保持平行为特征的未折叠蛋白质的普遍初始热力学亚稳态。该亚稳态使得未折叠的蛋白质能承载一定的扭矩和弯矩而不发生折叠，这避免了蛋白质发生错误折叠。三种可以破坏未折叠蛋白质氨基酸序列中特定氨基酸的热力学亚稳态的物理机制被指出，沿蛋白质主链传播的扭转波是启动这些物理机制的关键。通过研究这些物理机制可以破解蛋白质折叠的物理密码。这里我们揭示的未折叠蛋白质的普遍初始热力学亚稳态预示着蛋白质折叠问题

可能存在着以牛顿力学为基础的答案，蛋白质折叠问题可能是基础科学留给牛顿力学的遗产。

参考文献

[1] K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. Science 338, 1042-1046(2012).
[2] K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, The Protein Folding Problem. Annual review of biophysics 37, 289-316 (2008).
[3] Igor K. Lednev, Amyloid Fibrils: the Eighth Wonder of the World in Protein Folding and Aggregation. Biophysical Journal 106, 1433-1435 (2014).
[4] N. V. Grishin, Fold Change in Evolution of Protein Structures. Journal of Structural Biology 134, 167-185 (2001).
[5] C. B. Anfinsen, Principles that Govern the Folding of Protein Chains. Science 181, 223-230 (1973).
[6] J. C. Kendrew et al., Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. Nature 185, 422 (1960).
[7] K. Wakabayashi, [Accomplishment of Dr. Aaron Klug, winner of Nobel prize in chemistry, 1982]. Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme 28, 156-157 (1983).
[8] M. V. Rodnina, W. Wintermeyer, The ribosome goes Nobel. Trends in biochemical sciences 35, 1-5 (2010).
[9] a. G D Rose, R. Wolfenden, Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. Annual Review of Biophysics and Biomolecular Structure 22, 381-415 (1993).
[10] A. C. Clark, Protein folding: are we there yet? Archives of biochemistry and biophysics 469, 1-3 (2008).
[11] D. T. Clarke, A. J. Doig, B. J. Stapley, G. R. Jones, The α-helix folds on the millisecond time scale. Proceedings of the National Academy of Sciences of the United States of America 96, 7232-7237 (1999).
[12] E. H. L. Chen et al., Directly monitor protein rearrangement on a nanosecond-to-millisecond time-scale. Scientific Reports 7, 8691 (2017).
[13] E. Alm, D. Baker, Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proceedings of the National Academy of Sciences of the United States of America 96, 11305-11310 (1999).
[14] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank. Nature Structural Biology 10, 980 (2003).
[15] H. Berman, K. Henrick, H. Nakamura, J. L. Markley, The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Research 35, D301-D303 (2007).
[16] O. Collet, How does the first water shell fold proteins so fast? The Journal of Chemical Physics 134, 085107 (2011).
[17] C. M. Soares, V. H. Teixeira, A. M. Baptista, Protein Structure and Dynamics in Nonaqueous Solvents: Insights from Molecular Dynamics Simulation Studies. Biophysical Journal 84, 1628-1641 (2003).
[18] Brooks B R, Bruccoleri R E, Olafson B D, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations[J]. Journal of Computational Chemistry, 2010, 4(2):187-217.
[19] D. Eisenberg, Three-dimensional structure of membrane and surface proteins. Annual review of biochemistry 53, 595-623 (1984).
[20] Sussman J L, Abola E E, Lin D, et al. The protein data bank[J]. Genetica, 2000.